

Structured Prediction

מחר. אלקי

מסומנים מילים labels-מסומן יש מחר.

-מסומנים מילים לא מוסר

-מסומנים מילים מיוגים האפשריים מלמדה מיוגים אלה

Supervised מיוגים

Natural Language Parsing: מיוגים

אם: מיוגים

אם: מיוגים אלה האפשריים מיוגים, מיוגים מיוגים, מיוגים מיוגים

אם זה מיוגים מיוגים מיוגים מיוגים?

1. מיוגים מיוגים האפשריים מיוגים מיוגים

2. מיוגים מיוגים מיוגים מיוגים מיוגים

אם אסור מיוגים אלה מיוגים מיוגים אלה מיוגים מיוגים

מיוגים: מיוגים אלה מיוגים מיוגים מיוגים

אם זה מיוגים מיוגים! מיוגים מיוגים מיוגים מיוגים

מיוגים מיוגים מיוגים מיוגים

NER
POS
Sequence Prediction Problems: מיוגים מיוגים מיוגים

: Named Entity Recognition I

מיוגים: מיוגים מיוגים מיוגים מיוגים

מיוגים מיוגים מיוגים (Location, Company, ...)

אם מיוגים? מיוגים מיוגים מיוגים מיוגים

מיוגים מיוגים מיוגים

מיוגים מיוגים

1. מיוגים מיוגים מיוגים מיוגים מיוגים

מיוגים מיוגים מיוגים מיוגים מיוגים

2. מיוגים מיוגים מיוגים prediction מיוגים מיוגים

Part of Speech Tagging II

אם הדבר הזה נמצא בחקירה (אנחנו) המילים למינן (כיון).
אנחנו: נרצה להבין איך הם מילים שלא פסס לא האינו.
← נרצה להבין באינסטרומנטים של הדקטור.

איזה מילה יש לנו?

1. הסתברות של מילים: באילו הסתברות המילה saw

מופיעה בפסוק וקאילו הסתברות כלסצפ.

2. מילים שכלול: יש רצפים של חלקי משפט שיש מאוחר לא סקרים,

אם ניתן לשלול גיוגים בלי להסתכל על הוויזן שלה.

3. מאפיינים מורפולוגיים: $adverb \rightarrow 'ly'$

אפשר למצוא משהו על המילה מהמקום שלה.

סוג מילה
(אסוף)

הצורה אפשר להגיע מוצג, סקור, חפץ (אסוף) אם מסתכלים רק על

מילים בודדות. אלא אנחנו צולקים מה ביולוג הנורמלי ...

אנחנו נרצה להגדיר את המודל הסטטיסטי שלנו

הגדרה: סדרה (sequence) של מילים מקריים Y_1, \dots, Y_n

היא שרשרת מקרית (Markov Chain, מומאלי, מפרק ראשונה)

$$\forall i \quad P(Y_i | Y_1, \dots, Y_{i-1}) = P(Y_i | Y_{i-1})$$

$$(\Leftrightarrow) Y_i \perp \{Y_1, \dots, Y_{i-2}\} | Y_{i-1}$$

הסתברות משתנה
של Y_1, Y_2, \dots, Y_n

$$(\Leftrightarrow) P(Y_{1:n}) = \prod_{i=1}^n P(Y_i | Y_{i-1})$$

הנחה נוספת: נסמן את המילה הראשונה $Y_0 = \text{start}$ ק

למשל, אם זה כמו לא להקטין בשום דבר - רק/נחות.

הגדרה: 2 סדרות של מילים Y_1, \dots, Y_n , X_1, \dots, X_n הן HMM

(Hidden Markov Model) אם ההסתברות המשותפת שלהם מקיימת:

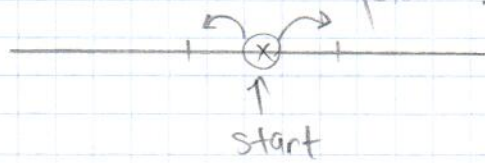
$$P(Y_{1:n}, X_{1:n}) = \underbrace{\prod_{i=1}^n P(Y_i | Y_{i-1})}_{P(Y_{1:n})} \cdot \underbrace{\prod_{i=1}^n P(X_i | Y_i)}_{P(X_{1:n} | Y_{1:n})}$$

↑ למטה מקיפים את Y הם שרשרת מקרית -

וזה סוג ההסתברות השותף -

$$P(Y_{1:n})$$

Sequence Prediction Problems



הערות: Y_i מיקור X_i מיקור

$$P(Y_0 = 0) = 1 \quad P(Y_i | Y_{i-1}) = \begin{cases} Y_{i-1} + 1 & \text{w.p. } \frac{1}{2} \\ Y_{i-1} - 1 & \text{w.p. } \frac{1}{2} \end{cases}$$

סומר $\{Y_i\}_{i=1}^n$ היא שרשרת מרקובית.

ונניח שסומר X_1, \dots, X_n שרשרת מיקור סגור למעלה או למטה, X_i מיקור.

$$P(X_i | Y_i) = \begin{cases} Y_i & \text{w.p. } \frac{1}{2} \\ Y_i + 1 & \text{w.p. } \frac{1}{4} \\ Y_i - 1 & \text{w.p. } \frac{1}{4} \end{cases}$$

הערות: 1. Y_i היא שרשרת מרקובית.

2. נניח X_i הוא מיקור סגור למעלה או למטה, X_i מיקור.

סומר X_i הוא מיקור סגור למעלה או למטה, X_i מיקור.

הערות: 1. Y_i היא שרשרת מרקובית.

2. נניח X_i הוא מיקור סגור למעלה או למטה, X_i מיקור.

סומר X_i הוא מיקור סגור למעלה או למטה, X_i מיקור.

נניח X_i הוא מיקור סגור למעלה או למטה, X_i מיקור.

POS

$X_{i:n}$ = המילים במשפט

$Y_{i:n}$ = המילים POS של המילים במשפט \rightarrow אנונימי

נניח שאנחנו $X_{i:n}, Y_{i:n}$ הם HMM.

האם ההנחות מתקיימות?

1. $Y_{i:n}$ היא שרשרת מרקובית: משמעות ההנחה סבה היא

שכל מילה "ציון" משמעותה מילים, מספיק לה "ציון" לכל מילה מילה.

למילים עוקבות במשפט. זה די הוליוודי כי המילים סתם קצת מקוונות.

אבל זה לא מילה מספיק \Rightarrow אפשר שרשרת מרקובית.

מספר של מילה או מילה יותר (מילה 2-מילה או יותר) אבל במובן

למילים סתם.

Sequence Prediction Problems

POS: קיום ההתחנה של HMM: התנאי

2. X_i כאי גלוי קלט פנימי חזר n - Y_i : התנה זו קדמה

לא מקבילת! כי אם אלה לא יוקעים מה התנה, אם

אם קדם מה התנה POS שלה, נקב אחרת עליה התנה מלאו תנאים.

התנאים של HMM:

1. transition probabilities: $t(y, y')$ - אסטרטגיה, מה ההסתברות שיהיה

$$t(y, y') = P(Y_i = y' | Y_{i-1} = y)$$

כמה? כמות:

$$\forall y' \sum_y t(y, y') = 1$$

וחיך להתקיים:

2. emission probabilities: $e(w, y)$ - ערך y אפשרי Y_i וזר w - X_i

$$e(w, y) = P(X_i = w | Y_i = y)$$

מה ההסתברות?

$$\forall y \sum_w e(w, y) = 1$$

ואם בן/נדרוש:

דוגמה: Bill saw that man yesterday

" " " " "

X_1 X_2 X_3 X_4 X_5

Name Verb Conj Noun Adverb

" " " " "

Y_1 Y_2 Y_3 Y_4 Y_5

$$P(X_{1:n}, Y_{1:n}) = t(\text{Name}, \text{start}) \cdot t(\text{verb}, \text{Name}) \cdot \dots$$

$$\cdot e(\text{Bill}, \text{Name}) \cdot e(\text{Saw}, \text{Verb}) \cdot \dots$$

ואם מה שגורו אחרת - זה אחרת אם e וזה אחרת אחר אחרת

סתומה סתומה סתומה סתומה סתומה

דוגמה: Sequence Prediction Problems

pos tags

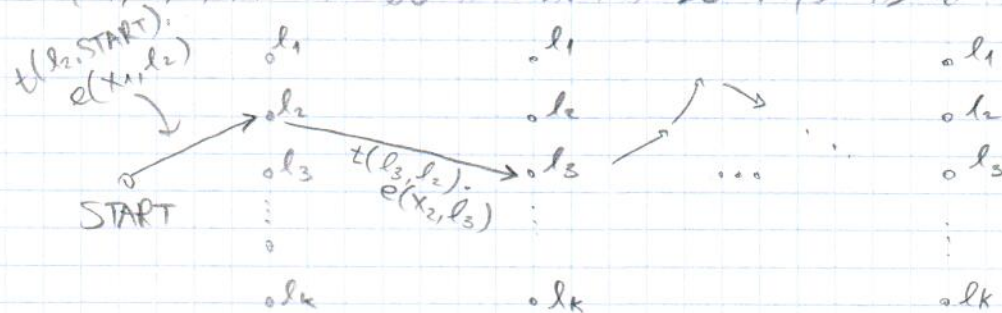
vocabulary : Inference: HMM → POS

$\{t(i,j)\}_{i,j \in L}$, $\{e(w,i)\}_{w \in V, i \in L}$: HMM נתון ונרצה למצוא את ה- $y_{1:n}$ עבור $x_{1:n}$
: עזרה, (עזרה) $x_{1:n}$: המילה

$$y^* = \underset{y_{1:n}}{\operatorname{argmax}} P(y_{1:n} | x_{1:n})$$

$$= \underset{y_{1:n}}{\operatorname{argmax}} P(y_{1:n} | x_{1:n}) \cdot P(x_{1:n})$$

$L = \{l_1, l_2, \dots, l_k\}$: המילים



$$w_{y_{i-1} \rightarrow y_i} = t(y_i, y_{i-1}) \cdot e(x_i, y_i)$$

יחסים ופרמטרים של המודל

: Viterbi Algorithm ? עזרה

$$\pi(t, j) = \max_{\substack{y_1, \dots, y_t \\ y_t = j}} P(y_{1:t}, x_{1:t})$$

$$\pi(t, j) = \max_{j'} \{ \pi(t-1, j') \cdot t(j, j') \cdot e(x_t, j) \}$$

כאשר t הוא מספר המילים
המקסימום של $\pi(t, j)$ הוא
המקסימום של $\pi(t-1, j')$

$$\pi(1, j) = t(j, \text{START}) \cdot e(x_1, j)$$

יש מקסימום אחד בלבד
ב- j

כאשר t הוא מספר המילים

למידה : Learning: HMM → POS

$$\ell(\{t(i,j)\}, \{e(w,i)\}) = \sum_{r=1}^M \sum_{k=1}^n [\log t(y_k^{(r)}, y_{k-1}^{(r)}) + \log e(x_k^{(r)}, y_k^{(r)})]$$

הפרמטרים של המודל, $\{x_{1:n}, y_{1:n}\}_{r=1}^M$

$$t(i,j) = n_{ij} / \sum_k n_{kj}, e(w,i) = n_{wi} / \sum_k n_{ki}$$

Maximum Entropy Markov Model : MEMM \rightarrow POS

הם מאפשרים מסתגלים למידה של HMM

שם סוגר 1 : HMM-הם יכולים להשתמש בכל features, כמו
 למשל מורפולוגיה של מילים. אך על מילים שלא הופיעו או
 הופיעו מראש מאובד - יש להם מאגר מילים.
 כי כל מילה נקראת במסגרת מסוימת של מילים.

הם מודל discriminative. כלומר - אין אפשרות לדקוד מה- x יש
 רק מהמילה x נחשב את ה- y .
 \Rightarrow כלומר צריך לתת את x .

$$P(y_{1:n} | x_{1:n}) = \prod_{i=1}^n P(y_i | y_{i-1}, x_{1:n}) = \quad (1)$$

conditional independence

$$= \prod_{i=1}^n \frac{e^{\langle \Phi(y_i, y_{i-1}, x_{1:n}, i), w \rangle}}{z(x_{1:n}, y_{i-1})}$$

כאשר Φ היא פונקציה ה-features של w ו- w יהיו הפרמטרים של המודל